

Research Article

Enriching Genomic Resources and Transcriptional Profile Analysis of *Miscanthus sinensis* under Drought Stress Based on RNA Sequencing

Gang Nie,¹ Linkai Huang,¹ Xiao Ma,¹ Zhongjie Ji,² Yajie Zhang,¹ Lu Tang,¹ and Xinquan Zhang¹

¹Department of Grassland Science, Animal Science and Technology College, Sichuan Agricultural University, Chengdu, Sichuan 611130, China

²Department of Agronomy, Purdue University, West Lafayette, IN 47906, USA

Correspondence should be addressed to Xinquan Zhang; zhangxq@sicau.edu.cn

Received 7 February 2017; Accepted 18 October 2017; Published 29 November 2017

Academic Editor: Graziano Pesole

Copyright © 2017 Gang Nie et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Miscanthus × giganteus is widely cultivated as a potential biofuel feedstock around the world; however, the narrow genetic basis and sterile characteristics have become a limitation for its utilization. As a progenitor of *M. × giganteus*, *M. sinensis* is widely distributed around East Asia providing well abiotic stress tolerance. To enrich the *M. sinensis* genomic databases and resources, we sequenced and annotated the transcriptome of *M. sinensis* by using an Illumina HiSeq 2000 platform. Approximately 316 million high-quality trimmed reads were generated from 349 million raw reads, and a total of 114,747 unigenes were obtained after de novo assembly. Furthermore, 95,897 (83.57%) unigenes were annotated to at least one database including NR, Swiss-Prot, KEGG, COG, GO, and NT, supporting that the sequences obtained were annotated properly. Differentially expressed gene analysis indicates that drought stress 15 days could be a critical period for *M. sinensis* response to drought stress. The high-throughput transcriptome sequencing of *M. sinensis* under drought stress has greatly enriched the current genomic available resources. The comparison of DEGs under different periods of drought stress identified a wealth of candidate genes involved in drought tolerance regulatory networks, which will facilitate further genetic improvement and molecular studies of the *M. sinensis*.

1. Introduction

The genus *Miscanthus* is a species of promising C4 perennial nonfood bioenergy grasses for cellulosic biofuel production [1]. Specifically, *Miscanthus × giganteus*, a hybrid generated from a cross between tetraploid *Miscanthus sacchariflorus* and diploid *Miscanthus sinensis*, has been intensively studied in Europe and North America as a biomass feedstock [2–7]. However, it is the only genotype currently available for use in most countries by its natural sterility and a narrow genetic base [8, 9]. Furthermore, it is highly risky and genetically difficult to improve *M. × giganteus* through breeding, posing limitations to its biomass productivity, abiotic stress tolerance, and climatic adaptation under some extreme conditions [10–12]. As a progenitor of *M. × giganteus*, *M. sinensis* was widely distributed around East Asia and it was

shown that abundant wild *M. sinensis* resources were distributed in China providing a comparable yield and well abiotic stress tolerance in some places [12–16].

Drought is a common environmental stress which induces adverse impacts on almost all aspects of plant development, growth, reproduction, and yield in a temperate area, and plants must adapt to this stress to survive [17, 18]. Plant drought tolerance is a complex quantitative trait, involving multiple pathways, regulatory networks, and cellular compartments [19]. Many of drought-induced or drought-repressed genes with diverse functions had been identified by molecular and genomic analysis in model plants. In *Arabidopsis*, 299 drought-inducible genes were identified through 7000 full-length cDNA microarray [20] and were classified into two groups including function proteins and regulatory proteins [21]. In rice, 73 dependable

TABLE 1: The quality report of *M. sinensis* RNA sample sequencing and unigene assembling under drought stress.

Samples	TCRs	Q20 (%)	TNU	MLU (nt)	N50	DC	DS
M0	55,128,976	97.19	92,444	931	1609	42,705	49,739
M1	54,603,020	97.23	90,148	883	1570	39,303	50,845
M2	52,019,400	97.30	88,455	830	1421	38,735	49,720
M3	51,366,992	97.20	81,159	846	1473	34,424	46,735
M4	51,745,196	97.26	89,320	852	1507	38,541	50,779
M5	51,337,262	97.29	88,416	855	1480	38,862	49,554
Total	316,200,846	/	114,747	1288	1854	65,203	49,544

TCRs: number of total clean reads; Q20: percentage of bases whose quality is larger than 20 in clean reads (%); TNU: total number of unigene; MLU: mean length of unigene (nt); DC: distinct clusters, which means that there are several unigenes wherein similarity between them is more than 70% in one cluster; DS: distinct singletons, which means a single unigene comes from a single gene.

genes were confirmed which were induced by drought, high salinity, or cold stress [22]. In a comparative analysis of 73 genes with those identified in *Arabidopsis*, 51 of them performed a similar function and revealed a considerable degree of similarity of drought stress response at the molecular level. Specially, genes involved in antioxidative metabolic pathways play an important role in detoxifying reactive oxygen species that can accumulate under drought stress conditions. In addition, it is well known that the phytohormone abscisic acid (ABA) level is essential for drought stress responses. Several genes involved in ABA biosynthesis and catabolism pathway in the drought stress responses have been identified in model plant and crops. In *Arabidopsis*, the 9-*cis*-epoxycarotenoid dioxygenase (NCED) family gene *AtNCED3* transcripts are rapidly induced by drought stress, which proved that it plays a crucial role in drought stress-inducible ABA biosynthesis [23]. CYP707A3 is strongly induced by rehydration after dehydration condition, which is a major enzyme for ABA catabolism in the drought stress response [24, 25]. Significantly, the introduction of many stress-inducible genes via gene transfer resulted in improved plant stress tolerance [25–28]. In consequence, discovering differential expression genes and analyzing the functions of these genes are important to further our understanding of the molecular mechanisms of plant drought stress response and tolerance regulation and ultimately facilitate the enhancement of plant drought tolerance through genetic manipulation.

It has been well known that many wild plants show high tolerant phenotypes against abiotic stresses, such as salt, drought, and oxidative stresses [29–31]. Based on the previous tests of drought and cold tolerance of *M. sinensis* in Europe, a much broader range of adaptation than *M. × giganteus* was found in this diploid species [10, 32], indicating that *M. sinensis* is considered possible to breed varieties with higher tolerance for frost and drought than *M. × giganteus* [14]. However, gene identification and molecular mechanisms involving in drought tolerance of *M. sinensis* are not well understood. Recently, high-throughput RNA sequencing technology was proved a powerful tool for gene discovering, gene expression, and physiological and biochemical metabolism realization under abiotic stress [33, 34]. In this study, we explored differential expression gene and transcriptional profiles of *M. sinensis* under different drought stress stages based on transcriptome analysis, which contributes to well

understand the change of regulatory mechanism processes and provide molecular bases for further revealing of metabolic networks associated with drought tolerance. The drought tolerance-related genes identified in this study aimed to provide varied candidate genes for *M. sinensis* genetic improvement and crop breeding.

2. Results and Discussion

2.1. Sequencing Analysis and De Novo Assembly. A total of 6 RNA samples from *M. sinensis* drought tolerance genotype “M2010228” were sequenced using the Illumina HiSeq 2000 platform, and 349,393,396 raw reads were generated. After filtering and trimming the raw reads, a total of 316,200,846 high-quality clean reads were assembled into 114,747 unigenes using Trinity [35]. The length of unigenes ranged from 200 to 12,275 nt, the average length of unigenes was 1288 nt, and the total N_{50} was 1854 nt. Compared to previous de novo assembly of *M. × giganteus* by using ABySS and Phrap, the contigs obtaining longer than 200 bp were greater in this study, and a contig N_{50} length was longer than that of 1459 bp. All the unigenes were divided into two classes by gene family clustering, with a total of 65,203 distinct clusters identified which contained several similar unigene sequences (more than 70%) in each cluster and the other total of 49,544 distinct singletons generated with single unigene (Table 1).

The Illumina HiSeq 2000 platform, a short-read-based technology [36], was used leading to the generation of large-scale genomic and transcriptomic data for nonmodel crops. Furthermore, high-throughput RNA sequencing (RNA-Seq) technologies are more accurate and sensitive for detecting both low and high levels of gene expression [37]. Generally, Illumina sequencing platform is more cost-effective when compared to Roche 454 sequencing [38] and has been successfully used in transcriptome sequencing of many plant species [39–42]. In this study, an average of 52,700,141 clean reads was got from each sample, in which the number of clean reads was greatly larger than that of sequencing based on the 454 platform [43], although the average length of unigene was shorter. Therefore, the results of this study not only provide additional valuable genomic resources for *M. sinensis* but also construct reference for the comparison between two sequencing methods, which could

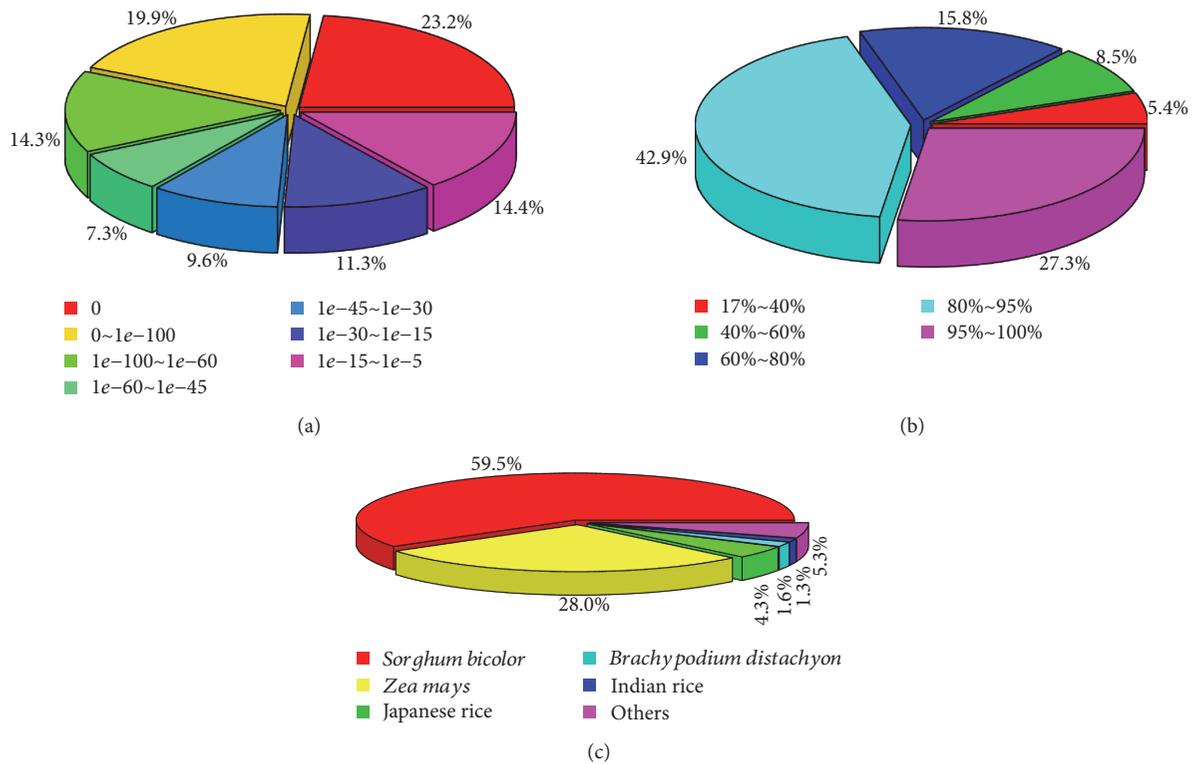


FIGURE 1: Unigene annotation results in the NR database. (a) displays the e value of the unigene annotation. (b) displays the identity of the similarity distribution. (c) displays the species distribution of annotated unigenes in NR database.

be used for further gene discovery and the identification of molecular markers.

2.2. Unigene Function Annotation and Classification Analysis.

Unigene annotation and classification provide rich information about expression profiles and predict the potential functions of the assembly. Furthermore, various databases for annotation could shed light on intracellular metabolic pathways and biological behaviors of genes. In this study, the 114,747 obtained unigene sequences from six *M. sinensis* samples were aligned to protein databases NR, Swiss-Prot, KEGG, COG, and GO by BLASTx and nucleotide database NT by BLASTn (e value $< 1e-5$). Amongst them, 91,894 (80.08%) unigenes had significant hits in NT database; 84,456 (73.60%) in NR; 64,145 (55.90%) in GO; 55,807 (48.63%) in Swiss-Prot; 55,557 (48.42%) in KEGG; and 38,458 (33.52%) in COG. In total, 95,897 (83.57%) unigenes were annotated using at least one database.

Within NR annotation, 74.3% of e value was $< 1e-30$ (Figure 1(a)) and 86% of similarity distribution was $> 60\%$ (Figure 1(b)). For the species distribution of all unigenes identified from six samples, the most frequent and significant annotation hits in the databases were matched to two well-annotated Poaceae plant species, including 59.5% of them which were annotated to *Sorghum bicolor* and 28.0% to *Zea mays* (Figure 1(c)). It is no surprise that the unigenes were annotated to these two species. Included within the Andropogoneae are major crops such as maize (*Zea mays* L.), sorghum (*Sorghum bicolor* L. Moench), and sugarcane (*Saccharum officinarum* L.) and species in the genus

Miscanthus. Domesticated and wild grass species in the Andropogoneae tribe are important sources of food, feed, fiber, and fuel [44]. Previous studies showed that the high utility of sorghum as a reference genome sequence for Andropogoneae grasses was widely used for *M. sinensis* in genome-wide association analysis and QTL mapping [16, 44–46]. Swaminathan et al. [47] constructed a framework genetic map of *M. sinensis* using single-nucleotide variant (SNV) markers which were developed by deep RNA sequencing and comparison with the genomes of sorghum maize and rice (*Oryza sativa*). Ma et al. [48] created high-resolution genetic mapping of *M. sinensis* and revealed that sorghum has the closest phylogenetic relationship to *Miscanthus* by comparing the genome sequences to several grass species. Besides, the similarity of *Miscanthus* transcripts to the gene models and ESTs of sorghum, sugarcane, maize, rice, and *Brachypodium distachyon* was assessed by Barling et al. [49] and showed that a large portion of similarity was contributed to sugarcane ESTs and sorghum gene models with most matches sharing over 95% identity. In this study, the results of unigene annotation by *M. sinensis* RNA-seq showed that most of the gene annotations (59.5%) were aligned to the sorghum database, which is consistent with previous studies with the high utility of sorghum as a reference genome sequence for genus *Miscanthus*, supporting that the sequences obtained in our study were annotated properly.

The Clusters of Orthologous Groups (COGs) of proteins were delineated by comparing protein sequences encoded in complete genomes, representing major phylogenetic lineages. Each COG consists of individual proteins or groups of

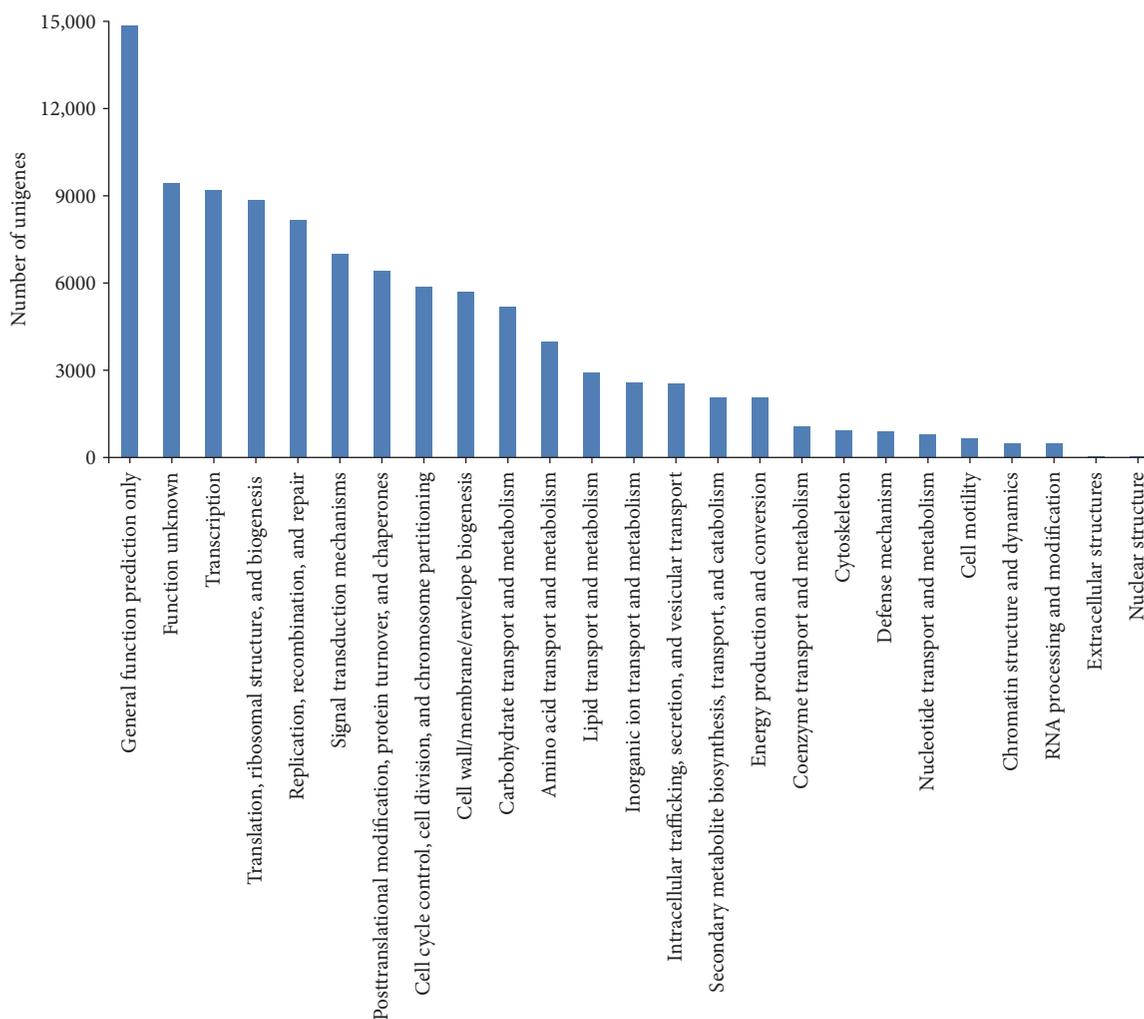


FIGURE 2: Function classification of all unigenes annotated in the COG database. A total of 38,458 unigenes were annotated under 25 function categories including (1) nuclear structure; (2) extracellular structures; (3) RNA processing and modification; (4) chromatin structure and dynamics; (5) cell motility; (6) nucleotide transport and metabolism; (7) defense mechanisms; (8) cytoskeleton; (9) coenzyme transport and metabolism; (10) energy production and conversion; (11) secondary metabolite biosynthesis, transport, and catabolism; (12) intracellular trafficking, secretion, and vesicular transport; (13) inorganic ion transport and metabolism; (14) lipid transport and metabolism; (15) amino acid transport and metabolism; (16) carbohydrate transport and metabolism; (17) cell wall/membrane/envelope biogenesis; (18) cell cycle control, cell division, and chromosome partitioning; (19) posttranslational modification, protein turnover, and chaperones; (20) signal transduction mechanisms; (21) replication, recombination, and repair; (22) translation, ribosomal structure, and biogenesis; (23) transcription; (24) function unknown; and (25) general function prediction only.

paralogs from at least 3 lineages and thus corresponds to an ancient conserved domain. The results of COG analysis indicated that 38,458 unigenes were annotated under 25 categories (Figure 2), among which were mainly classified into general function prediction (14,852 unigenes, 38.6%); transcription (9175 unigenes, 23.9%); translation, ribosomal structure, and biogenesis (8864 unigenes, 23.0%); replication, recombination, and repair (8162 unigenes, 21.2%); signal transduction mechanisms (6998 unigenes, 18.2%); and posttranslational modification, protein turnover, and chaperone function (6437 unigenes, 16.7%). In addition, there still have 9442 unigenes (24.6%) classified into unknown function category indicating that the unigenes identified from *M. sinensis* transcriptome under drought stresses were very different in biological functions involving in transport

and metabolism, cellular processes and signaling, and information storage and processing.

As an international standardized gene functional classification system, Gene Ontology (GO) offers a dynamic updated controlled vocabulary and a strictly defined concept to describe properties of genes and their products in any organism [50, 51]. In total, 64,145 (55.9%) unigenes were annotated to at least one of the three ontologies: molecular function, cellular component, and biological process in the GO database. In comparison, 58% of sorghum genes have GO annotation [52] as the most closely related plant to *M. sinensis*, indicating that our transcript assemblies afforded functional annotation of a comparable percentage of gene products to that of annotated plant species despite the current lack of a reference genome sequence.

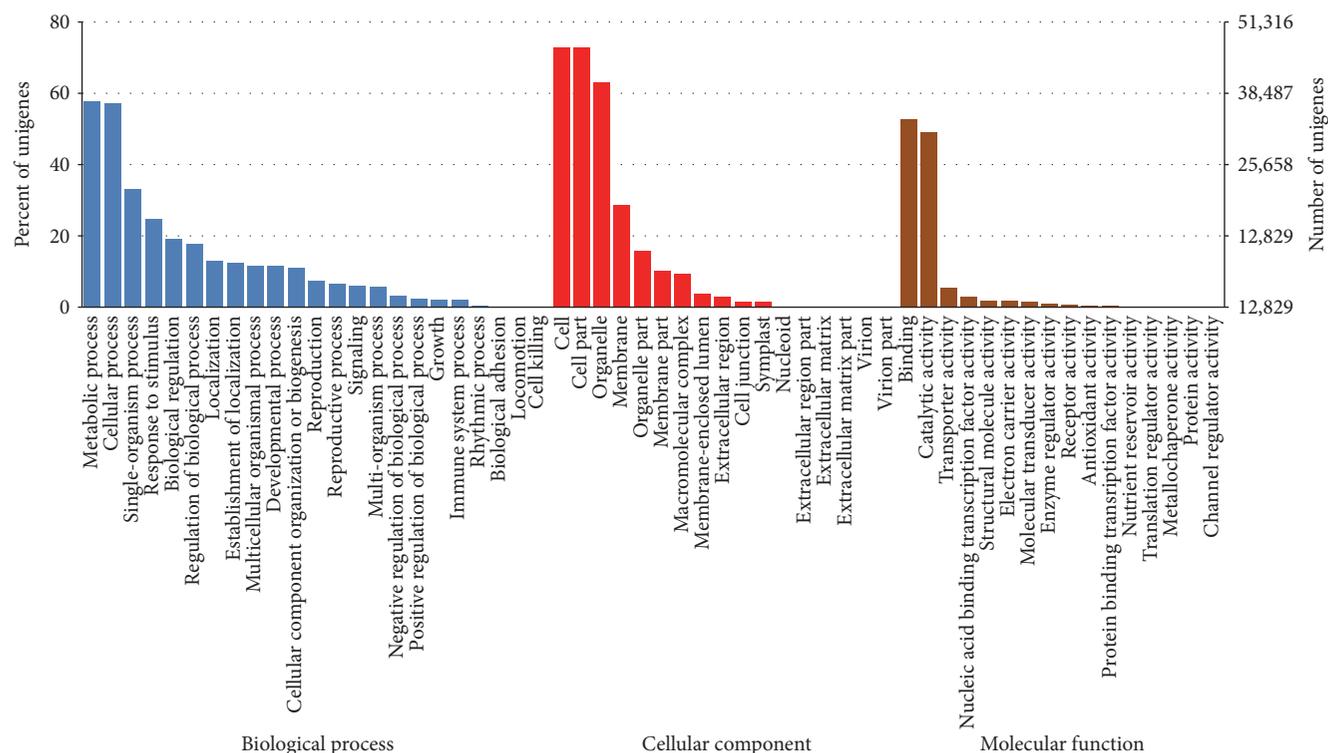


FIGURE 3: Function classification of all unigenes annotated in the GO database. A total of 64,145 unigenes were annotated under 23 different biological process categories, 17 cellular component categories, and 16 molecular function categories.

The distribution of all annotated unigenes in these three GO categories is shown in Figure 3. Among 23 different biological processes, metabolic process (57.9%), cellular process (57.0%), and single-organism process (33.2%) were the three most abundant GO categories responding to stimuli and biological regulation, suggesting active cellular and metabolic functions in *M. sinensis* leaves when exposed to drought stress. The frequent classes in cellular component were the cell part (72.9%), cell (72.9%), organelle (62.9%), and membrane (28.5%). Under the molecular function group, binding (52.5%) and catalytic activity (49.2%) were found to be the two mainly distributed categories as described, which are in agreement with the active metabolic functions in the examined tissues. With the help of GO functional classification, a large number of the unigenes were assigned to a diverse range of experimentally derived annotation. Our annotations provide a great foundation and a valuable resource for gene expression profile analysis, gene location, and gene isolation experiment in *Miscanthus* species. In addition, the main GO classifications identified through de novo transcriptome analyses in fundamental biological processes, cellular component, and molecular function were similar to previous reported studies in *M. sinensis*, *Sorghum bicolor*, [52] and *Hemarthria* [53], suggesting that our transcripts are the representative of a comprehensive *Miscanthus* transcriptome within the Andropogoneae tribe.

The networks of gene interactions in cells could be well understood by the KEGG pathway analysis. In this study, all the unigenes were analyzed in the KEGG pathway database and 55,557 unigenes were mapped to twenty main

categories including 128 different KEGG pathways (Figure 4). Most of the assigned genes were involved in a metabolism process (50,235, 90.4%), such as amino acid metabolism, carbohydrate metabolism, nucleotide metabolism, energy metabolism, lipid metabolism, and glycan biosynthesis and metabolism, suggesting a large number of genes induced by various metabolic activities under drought stress. Furthermore, a significant portion of unigenes was involved in the genetic information processing pathways (26,942, 48.5%), including transcription, translation, folding, sorting and degradation, replication, and repair. In addition, some of the unigenes were classified into organismal systems (4377, 7.9%), cellular process (4522, 8.1%), and environmental information processing (3625, 6.5%). The annotation of unigenes provides a large information base involved in drought tolerance process and plant drought response pathways, which serve an efficient guidance for future gene expression, gene network analyses, and regulatory metabolic network identification.

In order to conduct the prediction of protein coding region (CDs), all unigenes are firstly aligned by BLASTx (e value < 0.0001) to protein databases in the priority order of NR, Swiss-Prot, KEGG, and COG. Proteins with the highest ranks in BLAST results are taken to decide the coding region sequences of unigenes, and the coding region sequences are translated into amino sequences with the standard codon table. Unigenes that cannot be aligned to any database are scanned by EST-Scan, producing nucleotide sequence (5'→3') direction and amino sequence of the predicted coding region. In this study, a total of 84,633 (73.8%)

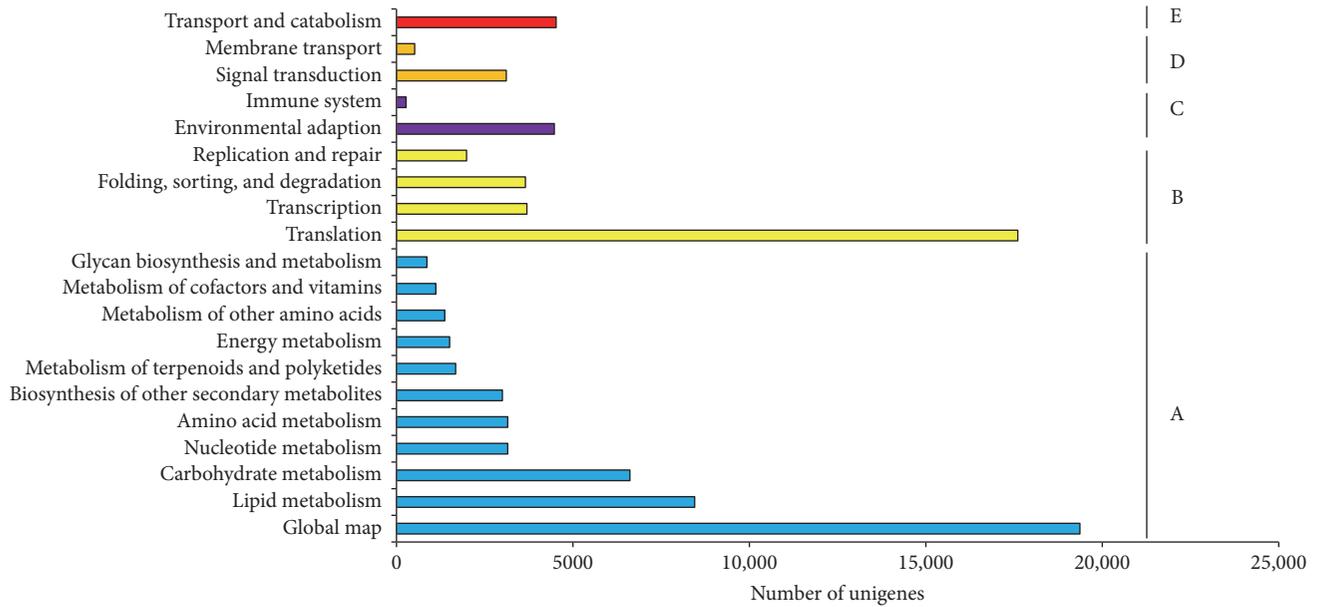


FIGURE 4: The KEGG pathway classification of all unigenes identified from six *M. sinensis* samples. A refers to metabolism category, B refers to genetic information processing category, C refers to organismal system category, D refers to environment information processing category, and E refers to cellular processes category.

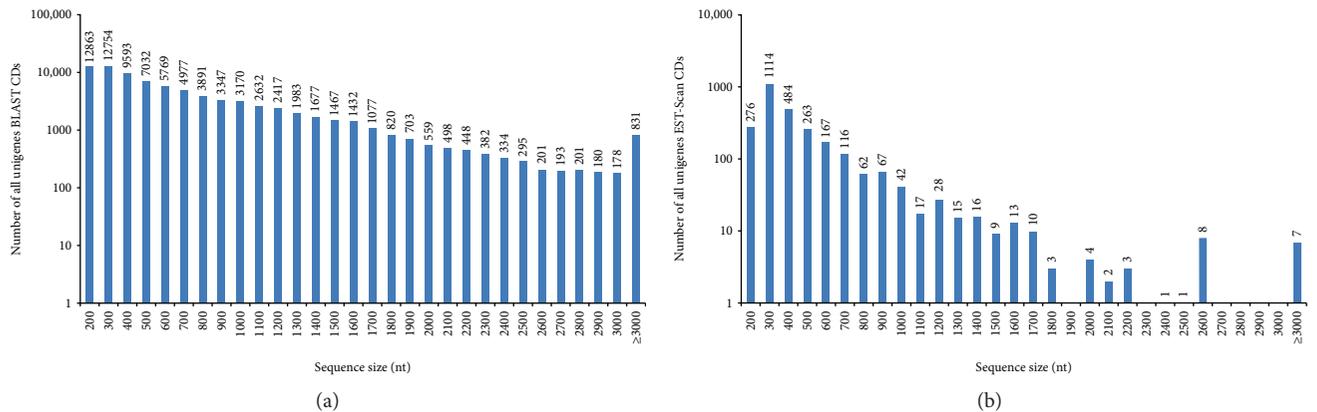


FIGURE 5: The length distribution of predicted CDs. (a) The length distribution of all unigenes BLAST CDs. (b) The length distribution of CDs scanned by EST-Scan.

unigenes were predicted to be CDs among the 114,747 assembled (Figure 5), of which 81,904 (96.8%) were aligned to the four previously discussed databases and another 2729 (3.3%) without BLAST hits were predicted by EST-Scan.

2.3. Differential Expression Gene Analysis. A total number of 5324 upregulated and 3276 downregulated differentially expressed genes (DEGs) were detected through the transcriptome comparison of well-watered (M0) versus drought stress 5 days (M1) in genotype “M20102208.” The numbers of DEGs found in M0 versus M2 were 3467 upregulated and 4175 downregulated; in M0 versus M3, 829 upregulated and 4683 downregulated; in M0 versus M4, 4370 upregulated and 3792 downregulated; and in M0 versus M5, 7536 upregulated and 5297 downregulated (Figure 6(a)).

The results showed that the number of upregulated DEGs was decreasing with the drought stress treatment time prolonging, and when the drought stress treatment continued 15 days, the upregulated DEGs just have 829. However, the number of upregulated DEGs was increased when the plant was exposed to drought stress 20 days and 30 days.

For getting the dynamic change profiles of DEGs of *M. sinensis* under drought stress, another group of comparison was conducted (Figure 6(b)). The results showed that drought stress treatments from 0 day (M0) to 5 days (M1), 5 days (M1) to 10 days (M2), and 15 days (M3) to 20 days (M4) were three critical steps with the largest change of the number of DEGs, while from 10 days (M2) to 15 days (M3) and 20 days (M4) to 30 days (M5) were relatively stable stages where the number of DEG is no obvious change. Especially

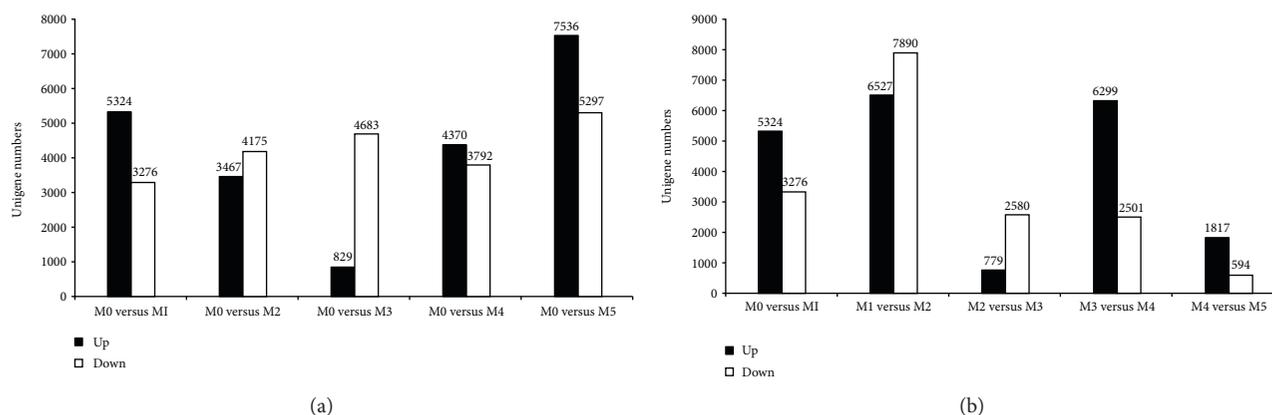


FIGURE 6: The number of DEGs at different time points under drought stress. (a) DEGs identified by the comparisons between well-watered (M0) and different drought stress treatments (M1, M2, M3, M4, and M5). (b) DEGs identified by the comparisons between two successive treatments.

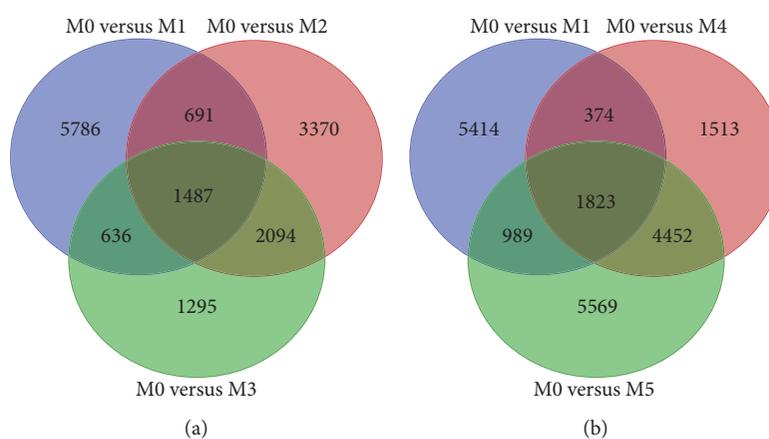


FIGURE 7: The similarity comparison of the DEGs between different time points under drought stress. (a) The DEGs identified from M0 versus M1, M0 versus M2, and M0 versus M3. (b) The DEGs identified from M0 versus M1, M0 versus M4, and M0 versus M5.

from 20 to 30 days, there just a total of 2411 DEGs were found indicating that *M. sinensis* exposed to drought stress 30 days almost had no molecular regulation or growth response and plant leaves showed obvious senescence phenotype.

The expression variation observed from RNA-seq provides a good representation of change in transcript profiles among samples [49]. Understanding the temporal expression change of regulated genes under drought stress may give insights into the gene related with drought stress adaptation in *M. sinensis*. After the similarity comparison of the DEGs identified from various stages of drought stress treatment, the DEGs from M0 versus M1 were highly dissimilar to other comparisons (just about 25% similar), while DEGs from M0 versus M3 have 65.0% matching to M0 versus M2, and the DEGs from M0 versus M4 have 76.9% similar to those from M0 versus M5 (Figure 7). The results were highly consistent with the previous results of dynamic change comparison in which the plants exposed to the 30-day drought stress treatment were divided into three main periods, that is, 0-day to 5-day slight stress, 5-day to 15-day medium stress, and 15-day to 30-day heavy stress.

At each defined phase, the differentially expressed genes are expected to have a specific function. More significantly, there were a substantial number of genes that exhibited modulated expression under drought stress. Interestingly, for the function analysis of the upregulated DEGs among medium stress periods, we found that after the 15-day drought stress treatment, the significantly high level expressed genes were mainly functions to cytochrome c oxidase, while stag-green related gene SORBIDRAFT [54] and were commonly highly expressed after the 10-day treatment. This indicates that a larger number of genes may participate in drought tolerance regulatory mechanisms and may be responsible for the plant phenotype under drought stress. In addition, various functions related to growth and development including putative ubiquitin carboxyl-terminal hydrolase superfamily, putative AP2/EREBP transcription factor superfamily, methyltransferase ZRP4, germin-like protein subfamily, lipoxygenase, and ATP synthase were also highly expressed within the initial slight stress and heavy stress period. These results indicate that drought stress 15 days could be a critical period for *M. sinensis* drought resistant molecular mechanism

TABLE 2: Identified assembled genes involved in ABA biosynthetic and catabolic pathway under drought stress.

Gene ID	Length	Swiss-Prot ID	Description	KO definition
Unigene18465	1664	sp O24592 NCED1_MAIZE	9-cis-epoxycarotenoid dioxygenase 1	EC: 1.13.11.51
Unigene2508	615	sp O24592 NCED1_MAIZE	9-cis-epoxycarotenoid dioxygenase 1	EC: 1.13.11.51
CL9892.Contig1	1319	sp Q05JG2 ABAH1_ORYSJ	Abscisic acid 8'-hydroxylase 1	EC: 1.14.13.93
CL8871.Contig2	1441	sp Q05JG2 ABAH1_ORYSJ	Abscisic acid 8'-hydroxylase 1	EC: 1.14.13.93
CL8871.Contig1	1635	sp Q05JG2 ABAH1_ORYSJ	Abscisic acid 8'-hydroxylase 1	EC: 1.14.13.93

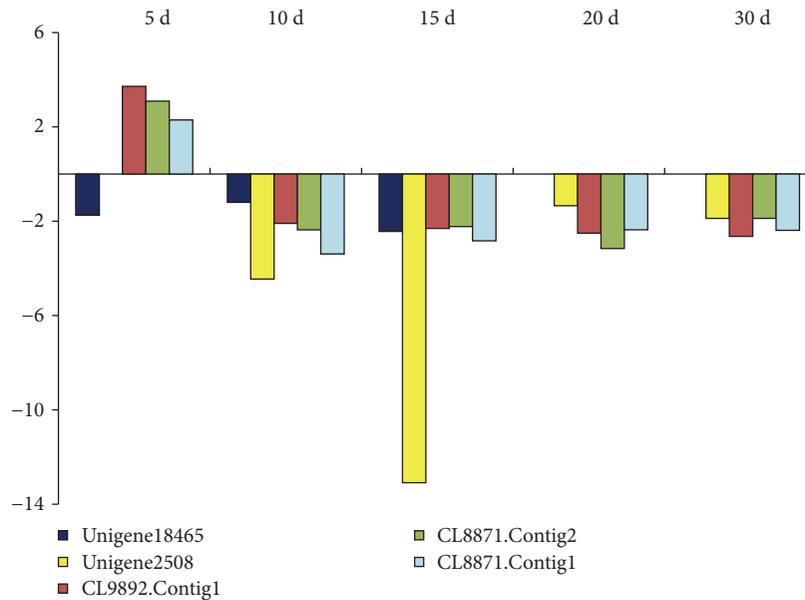


FIGURE 8: The differential expression level of key enzyme genes involved in ABA biosynthetic and catabolic pathway under drought stress. Unigene18465 and Unigene2508 functions were described as 9-cis-epoxycarotenoid dioxygenase 1. CL9892.Contig1, CL8871.Contig2, and CL8871.Contig1 functions were described as abscisic acid 8'-hydroxylase 1.

regulation, and when the relative soil water content reached 51.61% at 15 days, the chlorophyll of plant leaves begins to degrade to response to the stress.

In addition, the endogenous ABA level in plants is essential for various ABA-dependent stress responses, especially drought and salt stresses. Recently, with the molecular basis of ABA biosynthesis and catabolism established, the increasing concern about ABA biosynthetic and catabolic enzyme gene-expressed profiles is essential for the identification of drought stress response regulatory networks. Involved in carotenoid biosynthesis pathway, 9-cis-epoxycarotenoid dioxygenase (NCED) is a key enzyme for ABA biosynthesis, which was originally identified from maize viviparous 14 mutants playing a crucial role in drought stress-inducible genes [55]. On the other hand, abscisic acid 8'-hydroxylase (CYP707A) is an important enzyme gene for the ABA catabolism oxidative pathway in the drought stress response, which belongs to a class of cytochrome P450 monooxygenases [24]. In this study, we focused on the dynamic expression profile of these two DEG groups described above to identify the drought stress response mechanism of *M. sinensis* (Table 2). The results showed that the NCED gene

downregulated differential expression under drought stress in *M. sinensis* and the highest expressed at 15 days. The NCED transcripts are rapidly induced by drought stress to promote the ABA accumulation and to enhance drought tolerance of the plant [23]. However, among all the periods, we did not find the upregulated differential expression of NCED, indicating that the 5-day drought stress treatment in this study could be late for discovering the NCED gene in time, since the short-chain dehydrogenase/reductase (ABA2) was detected upregulated at 5 days. For CYP707A, the DEGs upregulated at 5 days and downregulated at the following periods represent that when the plant accepts the stress signal molecular, the ABA catabolism process was slow down for ABA accumulation to forcing the drought tolerance, and this could be an important mechanism for *M. sinensis* drought tolerance (Figure 8). Various crops could regulate drought stress response in different ways, although we indicate that *M. sinensis* drought tolerance was enhanced by the downregulated ABA catabolism pathway through the differential expression patterns; further studies are needed for function validation in knockout mutants or transgenic plants.

3. Experimental Design

3.1. Drought Treatment and Sample Collection. A wild drought tolerance genotype “M20102208” was collected from Sichuan Province (highway side, N 30° 08', E 103° 14'). All the plants were propagated through rhizome division from a single individual. The plants were planted in plastic pots (20 cm in diameter and 25 cm in height) with soil mixture (50% loam with 50% fine sandy). *M. sinensis* plants were grown in a growth chamber at 30°C/25°C, 16 h/8 h (day/night), 70% relative humidity, and 500 $\mu\text{mol photons m}^{-2}\cdot\text{s}^{-1}$. After three-month establishment, all the replications were subjected to nature drought stress treatment. Prior to treatment, all the pots were well-watered and the soil water content (SWC) was measured by the Soil Moisture Equipment TDR 300 (Santa Barbara, CA, USA). Naturally, water stress was applied by stopping water for 30 days. Leaf samples were collected and immediately frozen in liquid nitrogen from three replications at 0- (SWC=91.57%), 5- (SWC=85.55%), 10- (SWC=73.12%), 15- (SWC=51.61%), 20- (SWC=39.89%), and 30-day (SWC=26.41%) drought stress treatments for RNA extraction. Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen, USA) according to the manufacturer's instructions. RNA purity, concentration, and integrity were assessed using the RNA Nano 6000 Kit for the Agilent 2100 Bioanalyzer 2100 System (Agilent Technologies, USA). After RNA isolation and quality assessment, samples were stored at -80°C until the cDNA library construction and transcriptional assay were completed.

3.2. Library Construction and Sequencing. A total of 5 μg of total RNA per sample was used to construct the cDNA libraries. In all, six cDNA sequencing libraries of *M. sinensis* were constructed using the NEB Next® Ultra™ RNA Library Prep Kit for Illumina (New England Biolabs, USA). Initially, the total RNA was treated with RNase-free DNase I (NEB) for 30 min at 37°C and poly(A) mRNA was isolated from total RNA using poly-T oligo-linked magnetic beads. Following purification, the poly(A)-containing mRNA was fragmented into 200–250 bp pieces using fragment buffer (Ambion), and the first-strand cDNA was synthesized using random hexamer primers and the short fragments as templates. The products were then treated with RNase H, and second-strand cDNA was synthesized by DNA polymerase I (16°C for 2 h). Finally, NEBNext Adaptor with hairpin loop structure was ligated to the cDNA and the 3' ends of the DNA fragments were adenylated in preparation for hybridization. Subsequently, the cDNA fragments were purified and the quality of the library was evaluated using the Agilent Bioanalyzer 2100 system. The index-coded samples were clustered on a cBot System using the TruSeq PE Cluster kit v3-cBot-HS (Illumina). After generating the clusters, Illumina sequencing (paired-end technology in the Illumina HiSeq 2000 platform) of the six libraries was performed for RNA-seq analysis.

3.3. RNA Sequence Analysis and Drought-Induced Transcriptomic Changes. Raw reads (accession numbers

SRP095822 in the NCBI SRA database) from six libraries were transformed from sequencing-received image data. Raw reads were filtered to remove those with only adapters, low quality reads, and unknown or reads with less than 20 bp in length. Following the calculation of sequence duplication, Q20, and the GC content of the clean reads, the de novo assembly of RNA-seq was conducted using Trinity (<http://trinityrnaseq.github.io>), which is specific for high-throughput transcript assembly of RNA-Seq data without a reference genome [35]. The unigenes were generated by Trinity modules, and the processes of sequence splicing and redundancy removing were employed. The unigenes were annotated against the following protein databases: NR (nonredundant NCBI protein sequences), KOG/COGs (Clusters of Orthologous Groups of proteins), Swiss-Prot (a manually annotated and reviewed protein sequence database), and KEGG Ortholog database using BLASTx searches (e value < $1e-5$). Protein function information can be predicted from the annotation of the most similar protein in those databases. If the results of the databases conflicted with each other, a priority order of NR, Swiss-Prot, KEGG, and COG was followed when deciding the sequence direction of unigenes. The KEGG pathway database records networks of molecular interactions in the cells and variants of them specific to particular organisms. GO functional annotation was conducted with NR annotation which offers a dynamic updated controlled vocabulary and a strictly defined concept to comprehensively describe properties of genes using BLAST2GO program.

3.4. Differential Expression Genes and Pathway Analysis. The gene expression level was calculated by the number of uniquely mapped reads per kilobase of exon fragments per million mappable reads (FPKM) using Cufflinks (<http://cufflinks.cbcb.umd.edu/>). For genes with more than one alternative transcript, the longest transcript was selected to calculate the FPKM. With the expression level of each gene calculated, the differential expression analysis was conducted. The false discovery rate (FDR) as a statistical method was used to determine the threshold of the p value in multiple hypothesis testing, and for the analysis, a threshold of the $\text{FDR} \leq 0.001$ and an absolute value of \log_2 ratio ≥ 1 were used to judge the significance of the gene expression differences. Tool edgeR23 was used to identify significantly up- and downregulated genes on the read count values of genes. The differentially expressed genes (DEGs) were used for GO and KEGG enrichment analyses. First, all of the DEGs were blasted in the GO database (<http://www.geneontology.org/>) and the gene numbers were calculated for each GO term with GO-Term Finder version 0.86 (<http://search.cpan.org/dist/GO-TermFinder/>). GO terms were defined as significantly enriched GO terms in DEGs, if the corrected p value was ≤ 0.05 . Pathway enrichment analysis identifies significantly enriched metabolic pathways or signal transduction pathways in DEGs when compared with the whole genome background. Both GO terms and KEGG pathways with a Q value ≤ 0.05 are significantly enriched in DEGs.

4. Conclusions

High-throughput RNA sequencing technology was proved a powerful tool for gene discovering, gene expression, and physiological and biochemical metabolism realization under abiotic stress. By using the Illumina HiSeq 2000 platform to sequence *M. sinensis* under drought stress, approximately 316 million high-quality trimmed reads were generated from 349 million raw reads and a total of 114,747 unigenes were obtained after de novo assembly of the trimmed reads. Furthermore, 95,897 (83.57%) unigenes were annotated to at least one database including NR, Swiss-Prot, KEGG, COG, GO, and NT, and most of the annotations (59.5%) were aligned to the sorghum database, which is consistent with previous studies with the high utility of sorghum as a reference genome sequence for genus *Miscanthus*, supporting that the sequences obtained in our study were annotated properly. Differentially expressed gene analysis under different stress periods indicates that drought stress 15 days (soil water content reached 51.61%) could be a critical period for *M. sinensis* response to drought stress. *M. sinensis* plays an important role in improving the genetic base, abiotic stress tolerance, and climatic adaptation for this genus as a nonfood bioenergy crop. Hence, the transcriptome sequencing of *M. sinensis* reported here provides useful information for gene identification and greatly enriches the genomic available resources. The comparison of DEGs under different periods of drought stress allowed us to identify a wealth of candidate genes involved in drought tolerance regulatory networks, which will facilitate further advancements in genetic and molecular mechanisms with desired traits in further *M. sinensis* breeding programs.

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

Xinquan Zhang, Linkai Huang, and Xiao Ma conceived the project and designed the experiments. Gang Nie, Yajie Zhang, and Lu Tang performed the experiment. Gang Nie, Zhongjie Ji, and Linkai Huang wrote the paper. All authors discussed the results and commented on the manuscript.

Acknowledgments

This work was supported by the Earmarked Fund for the Modern Agro-Industry Technology Research System (no. CARS-34) and the Grassland Basic Resources Investigation Research in China (2017FY100602).

References

- [1] I. Lewandowski, J. M. O. Scurlock, E. Lindvall, and M. Christou, "The development and current status of perennial rhizomatous grasses as energy crops in the US and Europe," *Biomass and Bioenergy*, vol. 25, no. 4, pp. 335–361, 2003.
- [2] J. M. Greef, M. Deuter, C. Jung, and J. Schondelmaier, "Genetic diversity of European *Miscanthus* species revealed by AFLP fingerprinting," *Genetic Resources and Crop Evolution*, vol. 44, no. 2, pp. 185–195, 1997.
- [3] T. R. Hodkinson, M. W. Chase, M. D. Lledo, N. Salamin, and S. A. Renvoize, "Phylogenetics of *Miscanthus*, *Saccharum* and related genera (Saccharinae, Andropogoneae, Poaceae) based on DNA sequences from ITS nuclear ribosomal DNA and plastid *trnL* intron and *trnL-F* intergenic spacers," *Journal of Plant Research*, vol. 115, no. 5, pp. 381–392, 2002.
- [4] T. R. Hodkinson, M. W. Chase, C. Takahashi, I. J. Leitch, M. D. Bennett, and S. A. Renvoize, "The use of DNA sequencing (ITS and *trnL-F*), AFLP, and fluorescent in situ hybridization to study allopolyploid *Miscanthus* (Poaceae)," *American Journal of Botany*, vol. 89, no. 2, pp. 279–286, 2002.
- [5] E. A. Heaton, F. G. Dohleman, and S. P. Long, "Meeting US biofuel goals with less land: the potential of *Miscanthus*," *Global Change Biology*, vol. 14, no. 9, pp. 2000–2014, 2008.
- [6] E. A. Heaton, F. G. Dohleman, and S. P. Long, "Seasonal nitrogen dynamics of *Miscanthus* × *giganteus* and *Panicum virgatum*," *Global Change Biology Bioenergy*, vol. 1, no. 4, pp. 297–307, 2009.
- [7] A. Hastings, J. Clifton-Brown, M. Wattenbach, P. Mitchell, and P. Smith, "The development of MISCANFOR, a new *Miscanthus* crop growth model: towards more robust yield predictions under different climatic and soil conditions," *Global Change Biology Bioenergy*, vol. 1, no. 2, pp. 154–170, 2009.
- [8] A. Nishiwaki, A. Mizuguti, S. Kuwabara et al., "Discovery of natural *Miscanthus* (Poaceae) triploid plants in sympatric populations of *Miscanthus sacchariflorus* and *Miscanthus sinensis* in southern Japan," *American Journal of Botany*, vol. 98, no. 1, pp. 154–159, 2011.
- [9] M. S. Dwiyanti, J. R. Stewart, A. Nishiwaki, and T. Yamada, "Natural variation in *Miscanthus sinensis* seed germination under low temperatures," *Grassland Science*, vol. 60, pp. 194–198, 2014.
- [10] I. Lewandowski, J. C. Clifton-Brown, B. Andersson et al., "Environment and harvest time affects the combustion qualities of *Miscanthus* genotypes," *Agronomy Journal*, vol. 95, no. 5, pp. 1274–1280, 2003.
- [11] L. V. Clark, J. E. Brummer, K. Glowacka et al., "A footprint of past climate change on the diversity and population structure of *Miscanthus sinensis*," *Annals of Botany*, vol. 114, no. 1, pp. 97–107, 2014.
- [12] K. G. Anzoua, K. Suzuki, S. Fujita, Y. Toma, and T. Yamada, "Evaluation of morphological traits, winter survival and biomass potential in wild Japanese *Miscanthus sinensis* Anderss. populations in Northern Japan," *Grassland Science*, vol. 61, no. 2, pp. 83–91, 2015.
- [13] H. Zhao, B. Wang, J. He et al., "Genetic diversity and population structure of *Miscanthus sinensis* germplasm in China," *PLoS One*, vol. 8, no. 10, article e75672, 2013.
- [14] G. Nie, X. Q. Zhang, L. K. Huang et al., "Genetic variability and population structure of the potential bioenergy crop *Miscanthus sinensis* (Poaceae) in Southwest China based on SRAP markers," *Molecules*, vol. 19, no. 8, pp. 12881–12897, 2014.
- [15] G. Nie, L. Huang, X. Zhang et al., "Marker-trait association for biomass yield of potential bio-fuel feedstock *Miscanthus sinensis* from Southwest China," *Frontiers in Plant Science*, vol. 7, p. 802, 2016.

- [16] J. M. Gifford, W. B. Chae, K. Swaminathan, S. P. Moose, and J. A. Juvik, "Mapping the genome of *Miscanthus sinensis* for QTL associated with biomass productivity," *Global Change Biology Bioenergy*, vol. 7, no. 4, pp. 797–810, 2015.
- [17] M. Seki, T. Umezawa, K. Urano, and K. Shinozaki, "Regulatory metabolic networks in drought stress responses," *Current Opinion in Plant Biology*, vol. 10, no. 3, pp. 296–302, 2007.
- [18] K. Shinozaki and K. Yamaguchi-Shinozaki, "Gene networks involved in drought stress response and tolerance," *Journal of Experimental Botany*, vol. 58, no. 2, pp. 221–227, 2007.
- [19] X. Yu, G. Bai, S. Liu et al., "Association of candidate genes with drought tolerance traits in diverse perennial ryegrass accessions," *Journal of Experimental Botany*, vol. 64, no. 6, pp. 1537–1551, 2013.
- [20] M. Seki, M. Narusaka, J. Ishida et al., "Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray," *The Plant Journal*, vol. 31, no. 3, pp. 279–292, 2002.
- [21] K. Shinozaki, K. Yamaguchi-Shinozaki, and M. Seki, "Regulatory network of gene expression in the drought and cold stress responses," *Current Opinion in Plant Biology*, vol. 6, no. 5, pp. 410–417, 2003.
- [22] M. Rabbani, K. Maruyama, H. Abe et al., "Monitoring expression profiles of rice genes under cold, drought and high-salinity stresses, and abscisic acid application using cDNA microarray and RNA gel-blot analyses," *Plant Physiology*, vol. 133, no. 4, pp. 1755–1767, 2003.
- [23] S. Iuchi, M. Kobayashi, T. Taji et al., "Regulation of drought tolerance by gene manipulation of 9-*cis*-epoxycarotenoid dioxygenase, a key enzyme in abscisic acid biosynthesis in *Arabidopsis*," *The Plant Journal*, vol. 27, no. 4, pp. 325–333, 2001.
- [24] T. Kushiro, M. Okamoto, K. Nakabayashi et al., "The *Arabidopsis* cytochrome P450 CYP707A encodes ABA 8'-hydroxylases: key enzymes in ABA catabolism," *The EMBO Journal*, vol. 23, no. 7, pp. 1647–1656, 2004.
- [25] S. Saito, N. Hirai, C. Matsumoto et al., "Arabidopsis CYP707As encode (+)-abscisic acid 8'-hydroxylase, a key enzyme in the oxidative catabolism of abscisic acid," *Plant Physiology*, vol. 134, no. 4, pp. 1439–1449, 2004.
- [26] J. Zhang, R. Creelman, and J. Zhu, "From laboratory to field. Using information from *Arabidopsis* to engineer salt, cold, and drought tolerance in crops," *Plant Physiology*, vol. 135, no. 2, pp. 615–621, 2004.
- [27] F. Z. Wang, Q. B. Wang, S. Y. Kwon, S. S. Kwak, and W. A. Su, "Enhanced drought tolerance of transgenic rice plants expressing a pea manganese superoxide dismutase," *Journal of Plant Physiology*, vol. 162, no. 4, pp. 465–472, 2005.
- [28] T. Umezawa, M. Fujita, Y. Fujita, K. Yamaguchi-Shinozaki, and K. Shinozaki, "Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future," *Current Opinion in Biotechnology*, vol. 17, no. 2, pp. 113–122, 2006.
- [29] R. P. Ellis, B. P. Forster, D. Robinson et al., "Wild barley: a source of genes for crop improvement in the 21st century?," *Journal of Experimental Botany*, vol. 51, no. 342, pp. 9–17, 2000.
- [30] D. Bartels, "Targeting detoxification pathways: an efficient approach to obtain plants with multiple stress tolerance?," *Trends in Plant Science*, vol. 6, no. 7, pp. 284–286, 2001.
- [31] V. Mittova, M. Guy, M. Tal, and M. Volokita, "Salinity up-regulates the antioxidative system in root mitochondria and peroxisomes of the wild salt-tolerant tomato species *Lycopersicon pennellii*," *Journal of Experimental Botany*, vol. 55, no. 399, pp. 1105–1113, 2004.
- [32] J. C. Clifton-Brown, I. Lewandowski, B. Andersson et al., "Performance of 15 *Miscanthus* genotypes at five sites in Europe," *Agronomy Journal*, vol. 93, no. 5, pp. 1013–1019, 2001.
- [33] P. Lan, W. Li, and W. Schmidt, "Complementary proteome and transcriptome profiling in phosphate-deficient *Arabidopsis* roots reveals multiple levels of gene regulation," *Molecular & Cellular Proteomics*, vol. 11, no. 11, pp. 1156–1166, 2012.
- [34] L. Pan, X. Zhang, J. Wang et al., "Transcriptional profiles of drought-related genes in modulating metabolic processes and antioxidant defenses in *Lolium multiflorum*," *Frontiers in Plant Science*, vol. 7, p. 519, 2016.
- [35] M. G. Grabherr, B. J. Haas, M. Yassour et al., "Full-length transcriptome assembly from RNA-Seq data without a reference genome," *Nature Biotechnology*, vol. 29, no. 7, pp. 644–652, 2011.
- [36] M. Trick, Y. Long, J. Meng, and I. Bancroft, "Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing," *Plant Biotechnology Journal*, vol. 7, no. 4, pp. 334–346, 2009.
- [37] B. Wang, G. Guo, C. Wang et al., "Survey of the transcriptome of *Aspergillus oryzae* via massively parallel mRNA sequencing," *Nucleic Acids Research*, vol. 38, no. 15, pp. 5075–5087, 2010.
- [38] O. Morozova, M. Hirst, and M. A. Marra, "Applications of new sequencing technologies for transcriptome analysis," *Annual Review of Genomics and Human Genetics*, vol. 10, no. 1, pp. 135–151, 2009.
- [39] E. Mizrachi, C. A. Hefer, M. Ranik, F. Joubert, and A. A. Myburg, "De novo assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by Illumina mRNA-Seq," *BMC Genomics*, vol. 11, no. 1, p. 681, 2010.
- [40] R. Garg, R. Patel, A. Tyagi, and M. Jain, "De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification," *DNA Research*, vol. 18, no. 1, pp. 53–63, 2011.
- [41] D. Zou, X. Chen, and D. Zou, "Sequencing, de novo assembly, annotation and SSR and SNP detection of sabaigrass (*Eulaliopsis binata*) transcriptome," *Genomics*, vol. 102, no. 1, pp. 57–62, 2013.
- [42] S. Yates, M. Swain, M. Hegarty et al., "De novo assembly of red clover transcriptome based on RNA-Seq data provides insight into drought response, gene discovery and marker identification," *BMC Genomics*, vol. 15, no. 1, p. 453, 2014.
- [43] C. Kim, T. H. Lee, H. Guo et al., "Sequencing of transcriptomes from two *Miscanthus* species reveals functional specificity in rhizomes, and clarifies evolutionary relationships," *BMC Plant Biology*, vol. 14, no. 1, p. 134, 2014.
- [44] K. Swaminathan, M. S. Alabady, K. Varala et al., "Genomic and small RNA sequencing of *Miscanthus × giganteus* shows the utility of sorghum as a reference genome sequence for Andropogoneae grasses," *Genome Biology*, vol. 11, no. 2, article R12, 2010.
- [45] G. T. Slavov, R. Nipper, P. Robson et al., "Genome-wide association studies and prediction of 17 traits related to phenology, biomass and cell wall composition in the energy grass *Miscanthus sinensis*," *The New Phytologist*, vol. 201, no. 4, pp. 1227–1239, 2014.
- [46] S. Liu, L. V. Clark, K. Swaminathan, J. M. Gifford, J. A. Juvik, and E. J. Sacks, "High-density genetic map of *Miscanthus*

- sinensis* reveals inheritance of zebra stripe,” *Global Change Biology Bioenergy*, vol. 8, pp. 616–630, 2015.
- [47] K. Swaminathan, W. B. Chae, T. Mitros et al., “A framework genetic map for *Miscanthus sinensis* from RNAseq-based markers shows recent tetraploidy,” *BMC Genomics*, vol. 13, no. 1, p. 142, 2012.
- [48] X. F. Ma, E. Jensen, N. Alexandrov et al., “High resolution genetic mapping by genome sequencing reveals genome duplication and tetraploid genetic structure of the diploid *Miscanthus sinensis*,” *PLoS One*, vol. 7, no. 3, article e33821, 2012.
- [49] A. Barling, K. Swaminathan, T. Mitros et al., “A detailed gene expression study of the *Miscanthus* genus reveals changes in the transcriptome associated with the rejuvenation of spring rhizomes,” *BMC Genomics*, vol. 14, no. 1, p. 864, 2013.
- [50] M. A. Harris, J. Clark, A. Ireland et al., “The Gene Ontology (GO) database and informatics resource,” *Nucleic Acids Research*, vol. 32, Supplement 1, pp. D258–D261, 2004.
- [51] X. Tang, Y. Xiao, T. Lv et al., “High-throughput sequencing and *de novo* assembly of the *Isatis indigotica* transcriptome,” *PLoS One*, vol. 9, no. 9, article e102963, 2014.
- [52] P. Chouvarine, A. M. Cooksey, F. M. McCarthy et al., “Transcriptome-based differentiation of closely-related *Miscanthus* lines,” *PLoS One*, vol. 7, no. 1, article e29850, 2012.
- [53] X. Huang, H. D. Yan, X. Q. Zhang et al., “*De novo* transcriptome analysis and molecular marker development of two *Hemarthria* species,” *Frontiers in Plant Science*, vol. 7, p. 496, 2016.
- [54] G. N. Chaudhari and B. Fakrudin, “Candidate gene prediction and expression profiling of near isogenic lines (NILs) carrying stay-green QTLs in rabi sorghum,” *Journal of Plant Biochemistry and Biotechnology*, vol. 26, no. 1, pp. 64–72, 2017.
- [55] S. H. Schwartz, B. C. Tan, D. A. Gage, J. A. Zeevaart, and D. R. McCarty, “Specific oxidative cleavage of carotenoids by VP14 of maize,” *Science*, vol. 276, no. 5320, pp. 1872–1874, 1997.